

# *Python para Data Science*

## *Aprendizado de Máquina*



Prof. Jorge Zavaleta, Prof. Sergio Serra  
Prof. Macário Costa



PET-SI

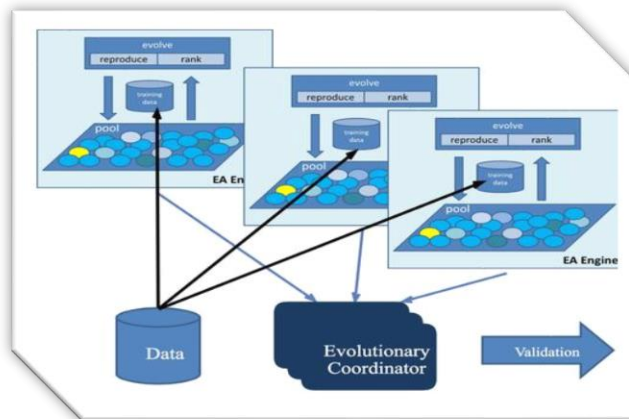


**UFRRJ**  
UNIVERSIDADE FEDERAL RURAL  
DO RIO DE JANEIRO



PET-SI

# Aprendizado de Máquina (ML)





PET-SI

## *Aprendizado de Máquina (ML)*

“Aprendizado de máquina é o campo de estudo que dá aos computadores a habilidade de aprender sem serem explicitamente programados”  
– Arthur Samuels, 1959.





PET-SI

# *Aprendizado de Máquina (ML)*

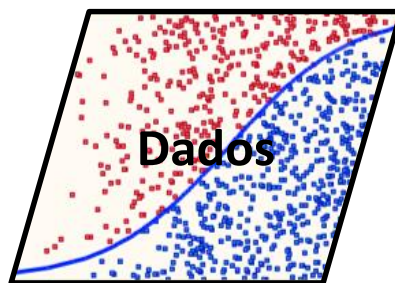
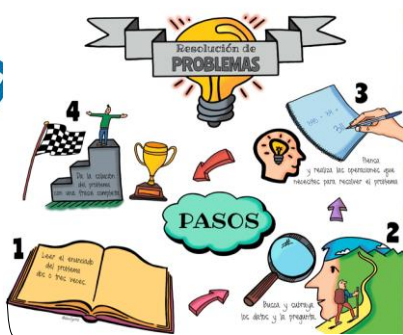
O que resolve  
ML?

- Problemas com grande número de regras ou processos manuais repetitivos.
- Problemas complexos sem boas soluções
- Obter soluções de grandes volumes de dados
- ...





PET-SI



Estudar o problema

Treinar Algoritmo ML

Avaliar solução

Analisar erros

executar

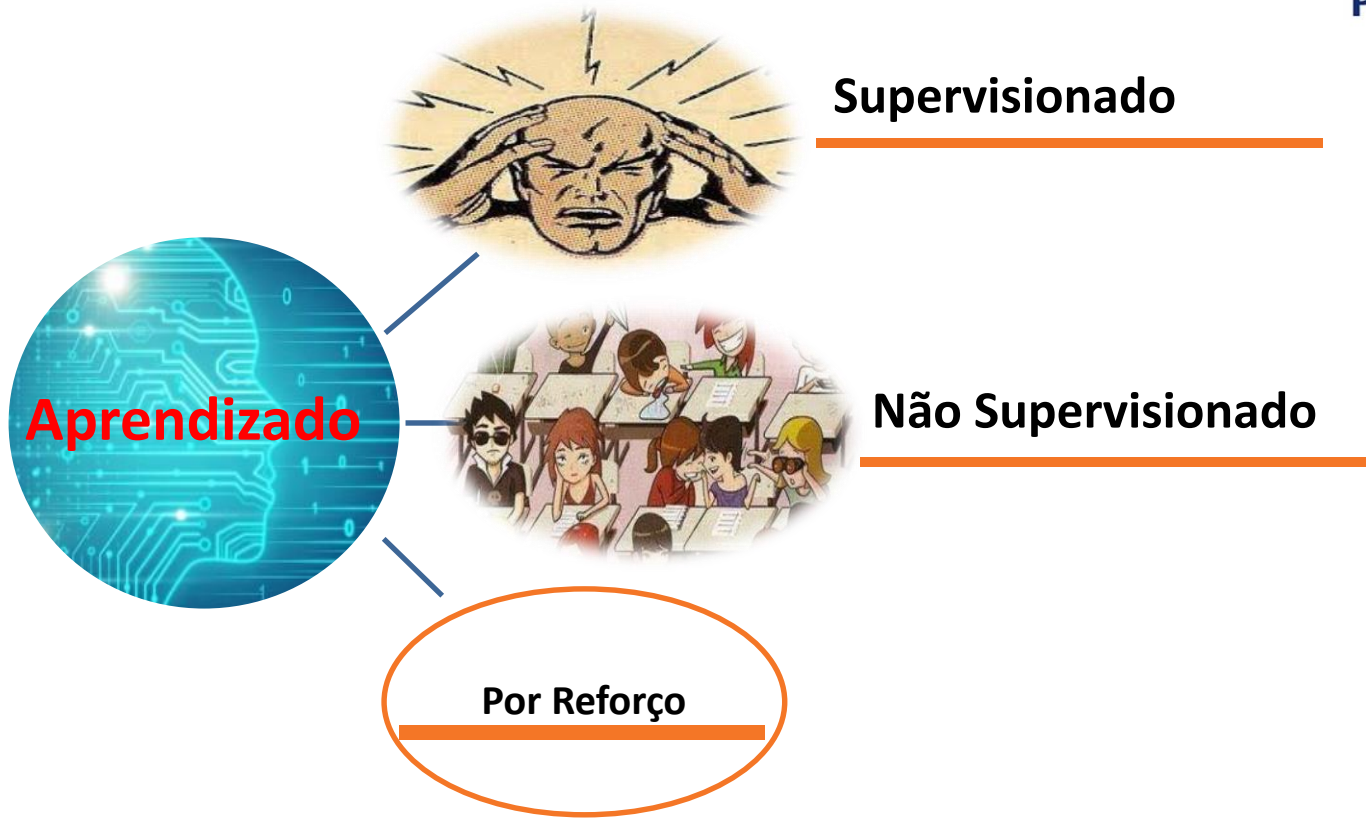






PET-SI

# ML – Tipos de Aprendizado

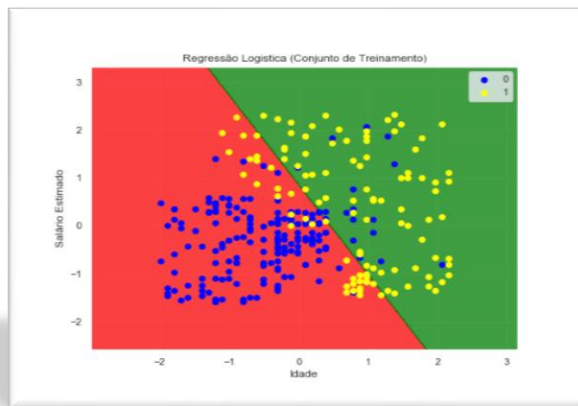




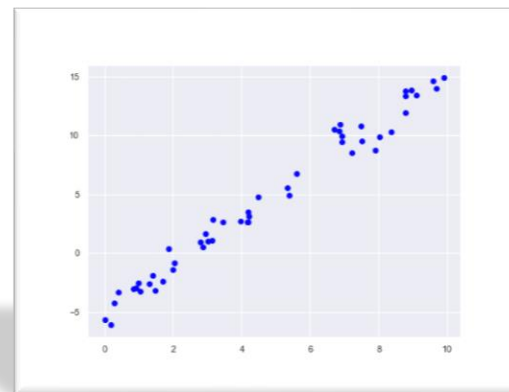
PET-SI

# Aprendizado Supervisionado

Os algoritmos de aprendizado supervisionados fazem previsões com base em um conjunto de exemplos de entrada. Neste tipo de aprendizagem existe um “professor/tutor” que avalia a resposta da rede ao padrão atual de entradas.



Classificação



Regressão

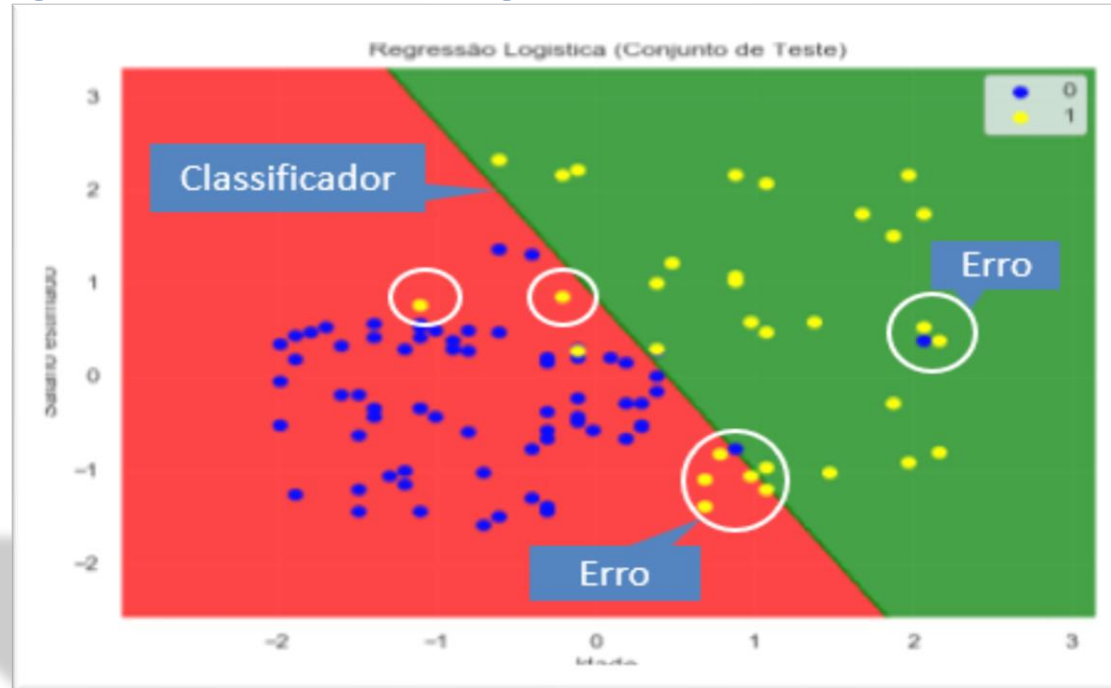




PET-SI

# Aprendizado Supervisionado

- **Classificação:** são métodos que buscam explicar uma variável categórica, com duas categorias (variável binária) ou mais.







PET-SI

## ***Classificação - Aplicações***

Filtros de spam, análise de sentimentos, detecção de fraude, segmentação de anúncios de clientes, previsões de rotatividade, sinalização de casos de suporte, personalização de conteúdo, segmentação de clientes, descoberta de eventos, genômica, eficácia de medicamentos, detecção de defeitos de fabrica etc.

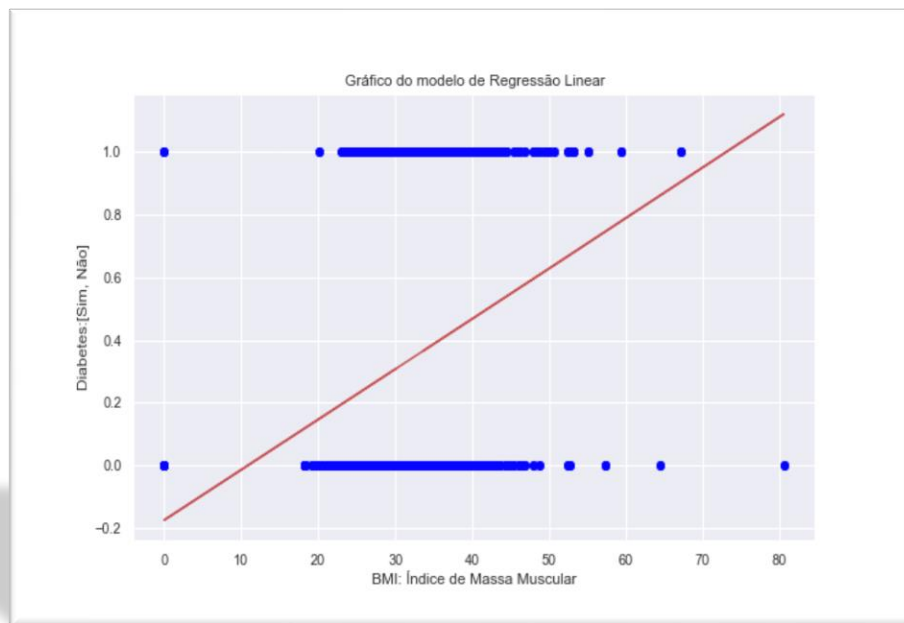




PET-SI

# Aprendizado Supervisionado

- **Regressão:** métodos de regressão buscam encontrar como uma variável evolui em relação a outras.





**PET-SI**

## ***Regressão***

- Previsão do mercado de ações,
- Previsão de demanda,
- Estimativa de preços,
- Otimização de lances de anúncios,
- Gerenciamento de riscos,
- Gerenciamento de ativos,
- Previsão do tempo (clima),
- Previsão de esportes

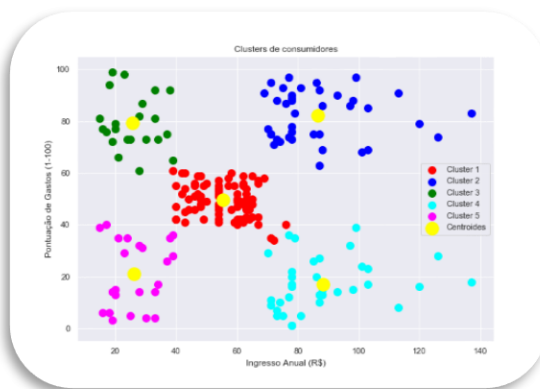




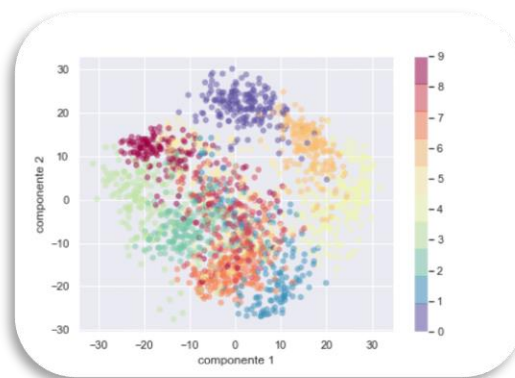
PET-SI

# Aprendizado Não Supervisionado

Não existe “professor/tutor”. A rede tem de descobrir sem ajuda as relações, padrões, regularidades ou categorias nos dados que lhe vão sendo apresentados e codificá-las nas saídas.



Clusterização



PCA

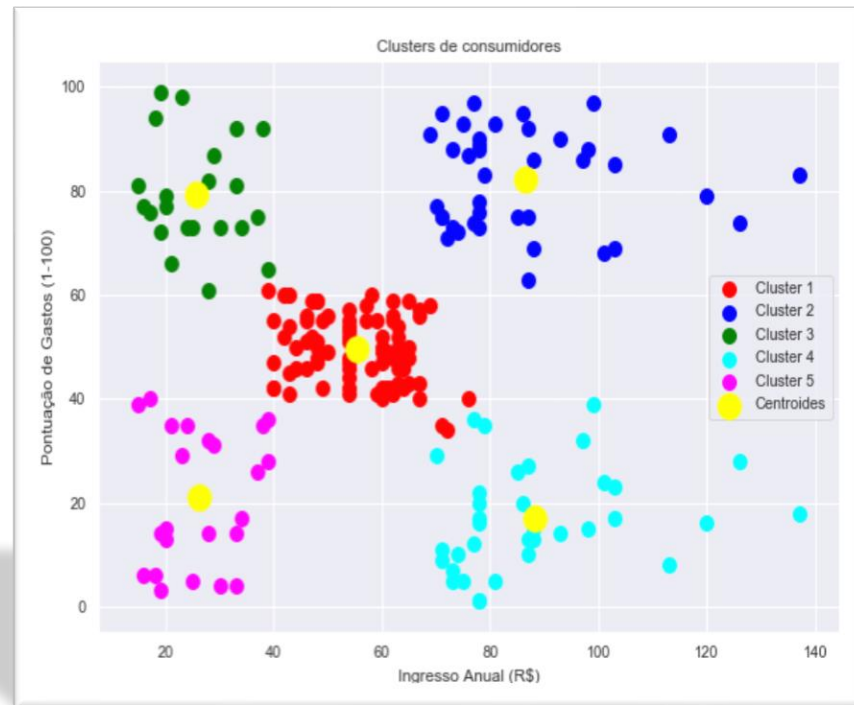




PET-SI

# Aprendizado Não Supervisionado

- **Clusterização** é a tarefa de dividir os pontos de dados em vários grupos com características semelhantes.
- Cada grupo possui um ponto central, denominado de **centroide**.



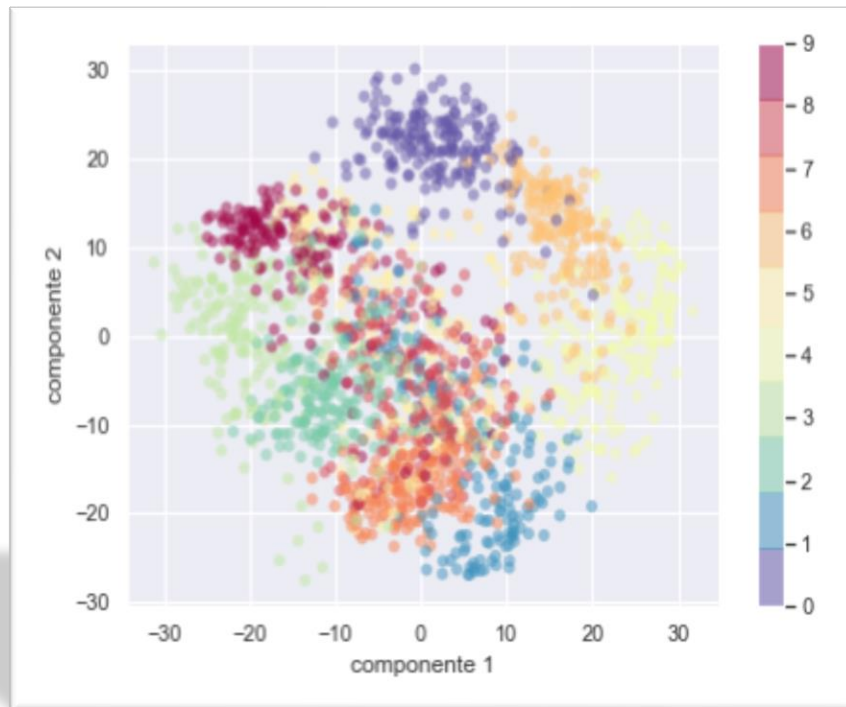




PET-SI

# Aprendizado Não Supervisionado

Principal Component Analysis (PCA -Análise de Componentes Principais) é um procedimento matemático que permite reduzir variáveis n-dimensionais em variáveis lineares.



# Matriz de Exemplos-Características

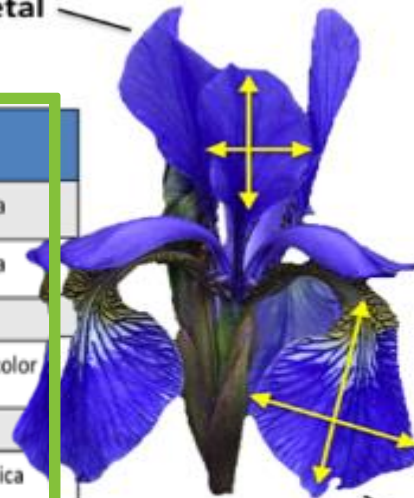
**Exemplos  
(instâncias, observações)**

	Sepal length	Sepal width	Petal length	Petal width	Class label
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3.0	1.4	0.2	Setosa
...					
50	6.4	3.5	4.5	1.2	Versicolor
...					
150	5.9	3.0	5.0	1.8	Virginica

**Características  
(atributos, medidas, dimensões)**

**Rótulo de classe  
(objetivo)**

Petal



Sepal



PET-SI

# *Prática Usando o Scikit-learn*



pandas  
 $y_i = \beta' x_{it} + \mu_i + \epsilon_{it}$

matplotlib

